

## REMARKS/ARGUMENTS

The Examiner objected to claim 6 for failing to include a period. Applicant amended claim 6 to correct this informality.

The Examiner rejected claims 1 - 13 under §102(e), as being as being anticipated by U.S. Patent No. 5,963,940 of Liddy et al. ("Liddy"). As discussed more fully and completely below, Liddy does not disclose or suggest a method or data structure for use in generating a text analysis program, as set forth in Applicant's claimed inventions. Therefore, Liddy cannot anticipate nor render obvious any of the pending claims.

Generally, the rejections are unwarranted, because Applicant claims a method, data structure and computer readable medium that are used to automatically generate a text analysis program (e.g., a natural language processing (NLP) system), while Liddy merely discloses an NLP system. The review by analogy equates a factory for building an airplane with an airplane. Although numerous patents exist for NLP systems, patents for systems and methods that automatically generate NLP systems are virtually unknown.

A text analyzer, also called an NLP system, parser, or information retrieval system in various embodiments, is a system that processes textual documents in order to extract usable information from the text. Complex representations such as parse trees, logical representations, and semantic structures are generated by a text analyzer in performing its functions. A text analyzer typically employs grammar rules, hierarchies of concepts, dictionaries, and other abstract representations as raw materials to aid in processing a text.

We have defined a text analyzer for three reasons: (1) to properly ground the following discussion, (2) to distinguish between a text analyzer and a system that generates a text analyzer, and (3) to clarify that the generation of representations by a text analyzer while processing a text document is distinct from the generation, merging, and generalization of rules and passes by a constructor system that automatically defines and builds a text analyzer.

The Examiner equates queries as described in Liddy with the text samples as described in present invention. Liddy describes a system in which a user may supply a query,

e.g., "sports Olympics" in order to find relevant documents in a collection of documents. Such query texts are *independent* of the document set to be searched. In contrast, Applicant discloses and claims text samples, which represent well-structured and meaningful units *within* the particular documents to be processed, and comprise "text that users have decided is a unit of interest, such as a name or idiomatic phrase." (Pending application at page 4, lines 19-20). For example, "Olympic event" might be a noun phrase within a particular document. These samples and the documents that they originate from serve as raw material for automatically generating a text analyzer. There can be no equivalency of purpose, organization, and use of the queries of Liddy with the text samples of the present invention. The text samples of the present invention could be viewed as "exemplars." One highlights and categorizes these exemplars within a document, then the present invention processes that same document in order to generate the rules of a text analyzer being constructed.

Another critical distinction in the methods of Liddy and Applicant should be noted. Liddy discloses methods to analyze single, isolated queries, whereas Applicant claims methods and a structure for collecting and organizing multiple related text samples into a hierarchy for the purpose of creating, merging, and generalizing rules and passes of a multi-pass text analyzer.

The Examiner further equates hierarchical information which may be built or used in Liddy with the organization of text samples into a hierarchy within the present invention. However, Liddy does not organize multiple queries in any fashion, whereas in the present invention a hierarchy is used to organize text samples for the purpose of generating a multi-pass text analyzer.

The present invention addresses the automatic generation of a multi-pass text analyzer by an iterative process. For example, a single-pass analyzer may be used to generate a second pass, then the two-pass analyzer may be applied so as to generate a third pass, and so on, until a complete text analysis system is created by repeatedly processing a hierarchy of samples and the documents containing those samples.

In contrast, the crux of Liddy is a particular embodiment of a text analyzer that attempts to rank documents based on their similarity to a single query text.

Thus, we find little overlap in goals, systems, methods, algorithms, inputs, and outputs between Liddy and the present invention. Indeed, the pending application cites several closely related patents that describe the automatic creation of text analysis capabilities from samples. Liddy is not related to this work, nor does Liddy cite these earlier precedents, due to the lack of relevance.

It is also the case that the present methods could be used to automatically generate parts of Liddy's system, if desired. Similarly, the present invention can be used to automatically generate virtually any of the patented text analyzers in the patent and academic literature. Applicant presents a novel way to create text analyzers, rather than manually writing rules and code. It also represents a new alternative to statistical methods for generating grammars from tagged and untagged document sets.

We now respond to the specific discussion of claims in the review. We quote relevant portions of the review and the cited portions of Liddy as needed. For convenience, we preserve the ordering of the Claims as presented in the review.

### ***Claims 1 and 13***

Because claims 1 and 13 include similar elements (a)-(d), the following discussion applies to both claims. The Examiner rejected independent claims 1 and 13, stating "Liddy discloses a natural language information retrieval system and method." Applicant reiterates that, in contrast, the claimed inventions disclose a method and computer readable medium for automatically generating a text analyzer or NLP system. Therefore, the claimed invention in its entirety is not disclosed or suggested in Liddy.

Furthermore, Liddy fails to disclose or suggest claim element (a): "providing a sample hierarchy" that includes "samples of text." The Examiner attempts to equate Liddy's queries, which the Examiner argues may "contain hierarchical information" with the claimed sample hierarchy. However, a query is merely a question, or a set of keywords, or a part of a document, or a whole document. As such, a query contains no hierarchical information per se. A query or multiple queries are not equivalent to a hierarchy of samples organized for the particular purpose of generating a text analyzer. Nor are queries, which are independent

texts, equivalent to the samples in the claimed invention, which are integrally connected to the documents containing those samples, which, taken together, serve as inputs for a system that generates a text analyzer. Therefore, Liddy fails to disclose or suggest Applicant's sample hierarchies or element (a) of the claimed inventions.

The Examiner further states that "generating one or more representations (rules or evidence sources) from the query" corresponds to Applicant's claim element (b): "extracting at least one rule from said sample hierarchy, said rule describing how to process a portion of text." However, Liddy's generation of representations by processing queries is merely the normal function of text analysis, for the purpose of extracting *content* from *text*. In contrast, Applicant's claim element converts a *sample* to a *rule* or pattern as part of defining a text analyzer.

We note here that although the Examiner suggests that Liddy may generate "rules" from queries, there is no such teaching in Liddy. Liddy does not claim to automatically generate rules, patterns, passes, or text analyzer definitions from queries, which is the essence of the pending claims. Therefore, Liddy further fails to disclose element (b) of the claimed invention.

The Examiner further asserts that Liddy's "generating from a query logical representations and evidence sources" is equivalent to claim element (c): "generating a pass from said rule, said pass containing instructions to operate a text analyzer." However, generating representations from a query is the enterprise of a text analyzer, and is entirely different from generating rules, passes, and a text analyzer, which is the enterprise of a factory for automatically constructing a text analyzer.

The Examiner further asserts that, in Liddy, "using the evidence sources the matcher scores documents", corresponds to claim element (d): "constructing a text analyzer containing said pass." Here the Examiner introduces a previously unmentioned system component, Liddy's matcher, which attempts to provide a score for the similarity of a query with a document. Such a system for ranking documents based on relevance or closeness to a query is not equivalent to a system that constructs a text analyzer. In fact, these processes are

entirely unrelated. Therefore, Liddy further fails to disclose or suggest element (d) of the claimed inventions.

Since Liddy fails to disclose or suggest any of claim elements (a)-(d), it cannot anticipate nor render obvious claim 1 or 13, or any claims depending from those claims (i.e., claims 2-11). For at least these reasons, Applicant respectfully requests that the related rejections be withdrawn.

***Claim 2***

Because claim 2 depends from claim 1, it cannot be anticipated or rendered obvious for at least the reasons set forth above. Additionally, claim 2 is patentable on the following independent basis. With respect to claim 2, the Examiner states "Liddy teaches that multiple operations are performed on a query to generation evidence sources and logical representations," and that this corresponds to Applicant's claim limitation that "said rule is generalized into multiple rules and multiple passes." However, performing a text analysis on a query, which may involve "multiple" operations, is not the same as algorithms and methods for collecting, merging, and generalizing rules to create multiple rules and multiple passes of a text analyzer.

Applicant reiterates that Liddy's method does not create rules, passes, or text analyzers. Rather, Liddy discloses a text analysis system.

***Claim 3***

Because claim 3 depends from claim 1, it cannot be anticipated or rendered obvious for at least the reasons set forth above. Additionally, claim 3 is patentable on the following independent basis. With respect to claim 3, the Examiner states that "Liddy teaches that five sources of evidence are used to compute five individual measures of similarity," and that this corresponds to Applicant's claim limitation "wherein multiple passes are added to said text analyzer." However, there is no relationship between using five measures of the content of queries and documents for assessing their similarity, on the one hand, and a system that generates the multiple passes of a text analyzer, on the other hand.

***Claim 4***

Because claim 4 depends from claims 1 and 3, it cannot be anticipated or rendered obvious for at least the reasons set forth above. Additionally, claim 3 is patentable on the following independent basis. The Examiner stated:

Liddy teaches that five individual measures of similarity are performed and their scores combined (col. 22, lns. 1-7), which corresponds to "said multiple passes are arranged in a cascading manner having a sequence of passes such that rules associated with a pass are applied to subsequent passes."

However, there is no equality between a system for scoring similarity of documents and queries to a method that builds a text analyzer with multiple passes, cascaded so that the rules built for a pass, along with all prior passes, serve to automatically construct the subsequent passes of an analyzer.

***Claim 5***

Because claim 5 depends from claim 1, it cannot be anticipated or rendered obvious for at least the reasons set forth above. Additionally, claim 5 is patentable on the following independent basis. With respect to claim 5, the Examiner states that "Liddy teaches the use of a tree-form logical representation of a query statement that is used during matching and scoring," which is to correspond to Applicant's claim limitation of "wherein the samples are associated with offset values, said offset values identifying locations in a parse tree data structure, said parse tree containing concepts stored at locations identified by said offsets."

Performing a text analysis on a query will produce a representation such as a "tree-form logical representation," which does **not** correspond to correlating offsets of a sample within a parse tree data structure. Here, Applicants claim a method for parsing documents and locating the samples contained within, by their offsets. In Liddy's method, the "document" and the query are always equivalent, that is, a query is not a part of a larger document. Hence, Liddy's methods have no bearing on Claim 5.

***Claim 6***

Because claim 6 depends from claims 1 and 4, it cannot be anticipated or rendered obvious for at least the reasons set forth above. Additionally, claim 6 is patentable on the following independent basis. With respect to claim 6, the Examiner asserts that "Liddy teaches that the user is given the opportunity to modify the analysis of the query," which is to correspond to Applicant's claim limitation of "allowing a user to control the extraction of rules from the sample hierarchy."

However, the relevant sections from Liddy (col. 28, lns. 16-30, Fig. 14A, Fig. 14B) disclose methods for modifying the output of a query. The user control disclosed in Applicant's claimed invention applies to modifying the algorithms that merge, generalize, and produce the rules and passes of a multi-pass text analyzer. What the user controls in Liddy as opposed to Applicant's invention cannot be equated.

***Claim 7***

Because claim 6 depends from claims 1 and 4, it cannot be anticipated or rendered obvious for at least the reasons set forth above. Additionally, claim 7 is patentable on the following independent basis. With respect to claim 7, the Examiner asserts that "Liddy teaches that the user can modify the representation of the query," which is to correspond to Applicant's claim limitation of "allowing a user to designate the properties associated with said properties controlling rule generation for a portion of the sample hierarchy."

However, Applicant claims a method that modifies attributes that will affect the way rules and passes of a multi-pass analyzer are generated, while Liddy discloses methods for altering the output produced by a text analyzer. The two methods cannot be equated.

***Claim 8***

Because claim 8 depends from claims 5 and 1, it cannot be anticipated or rendered obvious for at least the reasons set forth above. Additionally, claim 8 is patentable on the following independent basis. With respect to claim 8, the Examiner asserts, "Liddy teaches that the tree-form of the logical representation is used during scoring," which is to correspond to Applicant's claim limitation that "said concepts are retrieved from said parse tree and processed to form said rule."

However, Liddy's method performs text analysis on a query to produce representations that will then be used to rank relevant documents, while Applicant's method automatically selects concepts from a parse tree based on the offsets of samples within that parse tree, as part of an algorithm for collecting, merging, and generalizing rules based on a set of samples. These methods of Liddy and Applicant are unrelated.

***Claim 9***

Because claim 9 depends from claims 6, 4 and 1, it cannot be anticipated or rendered obvious for at least the reasons set forth above. Additionally, claim 9 is patentable on the following independent basis. With respect to claim 9, the Examiner asserts "Liddy teaches that the user can modify representations including proper noun and time frame," which is to correspond to Applicant's limitation of "allowing a user to designate attributes associated with said rules, said attributes guiding the application of said rules."

However, Meyers discloses methods for modifying attributes that will affect the way rules and passes of a multi-pass analyzer are generated, while Liddy discloses methods for altering the output produced by a text analyzer.

***Claim 10***

Because claim 10 depends from claim 1, it cannot be anticipated or rendered obvious for at least the reasons set forth above. Additionally, claim 10 is patentable on the following independent basis. With respect to claim 10, the Examiner asserts that "Liddy teaches that an alternative representation can encompass any or all of a plurality of representations," and that this corresponds to Applicant's claim limitation that "multiple rules are generalized and merged into a single rule if there is a difference between the multiple rules."

However, the relevant section of Liddy (col. 7, lns. 12-20) merely refers to a collection of representations produced during the text analysis of a single query. In contrast, Applicant discloses a method for merging and generalizing multiple rules from related samples during the creation of a text analyzer. These methods bear no relation to each other, nor does Liddy disclose methods for merging and generalizing multiple representations from



a single query, nor does Liddy disclose methods for merging and generalizing representations from multiple queries.

***Claim 11***

Because claim 11 depends from claims 10 and 1, it cannot be anticipated or rendered obvious for at least the reasons set forth above. Additionally, claim 11 is patentable on the following independent basis. With respect to claim 11, the Examiner asserts that "Liddy teaches that a query of any length is entered on the query screen 340 and that a query can be saved (inherently in a file) and recalled (from a file)," which is to correspond to Applicant's claim limitation that "samples may be contained in a sample file."

While the review introduces the notion of a file for queries, Liddy does not describe a file-based organization for queries. Nor should an isolated query be equated with a file that organizes multiple related text samples.

***Claim 12***

The Examiner rejected independent claim 12, stating "Liddy discloses a natural language information retrieval system and method." The Examiner asserts "Liddy's system includes the following: the ability to save previous requests (queries) where the text can be represented in a tree structure," which is to correspond to Applicant's claim limitation of an index for storing "samples comprising portions of text." However, Liddy's method does not organize the query texts themselves in a tree structure, but rather invokes a text analyzer to create representations such as tree structures from queries. In contrast, Applicant organizes the text samples themselves within a hierarchical structure. Liddy's queries are isolated texts, whereas Applicant's text samples are contained within full documents that are processed for the purpose of building a text analyzer.

The Examiner further asserts that Liddy's "generating one or more representations (rules or evidence sources) from the query where the representation(s) is used when matching documents in the database," corresponds to Applicant's limitation of "said samples used to generate rules for identifying patterns appearing in text." We reiterate here that generating representations of content is the standard function of a text analyzer, and cannot be equated

with methods that generate, merge, and generalize rules for the purpose of automatically creating a text analyzer. We note also that the review introduces the term “rules” to describe the output of Liddy’s system, while Liddy makes no such claims. Liddy does not disclose methods for generating grammar rules, patterns, or similar representations, other than standard text analyzer outputs such as logical representations, tree forms, and “sources of evidence.”

The Examiner asserts that Liddy’s “generating conceptual and term-based representations of each query used during matching,” corresponds to Applicant’s claim limitation of “said samples used to derive information from said identified patterns, said rules generated by parsing said text samples.” However, generating rules for the purpose of merging and generalizing them to create the multiple passes of a text analyzer cannot be equated merely with a text analysis method that outputs representations for the content of a query. Liddy’s method represents the content of a single, isolated query in order to find similar documents. In contrast, the claimed method can process a set of documents containing an embedded collection of organized, related text samples, in order to create, merge, and generalize rules and passes that automatically create a text analyzer.

The Examiner further asserts that Liddy’s “matching the query to indexed documents in multiple steps during scoring,” corresponds to Applicant’s claim limitation of an “index organized such that passes comprising operational steps and rules are generated in an order wherein simple patterns are recognized by said text analyzer.” Matching a query to multiple documents cannot be equated with generating a text analyzer comprising multiple passes (or “steps”). Although the Examiner asserts that Liddy’s method includes “multiple steps,” this does not equate to Applicant’s claimed method for generating a text analyzer with multiple passes, in which each pass consists of a set of rules.

The Examiner further asserts that Liddy’s “matching of evidence sources against a document starting with simple terms and moving to more complex structures such as paragraphs” corresponds to Applicant’s claim limitation that “said recognized simple patterns are used by said text analyzer and used to iteratively recognize more complex patterns.” Here Liddy discloses a method for scoring or ranking the similarity of a query to a set of

documents, using metrics of varying complexity. There is no notion in Liddy's method of using more complex methods that build upon simpler methods. "Moving to more complex structures" is not disclosed by Liddy. In any case, this does not equate to Applicant's method of generating more complex rules by building upon simpler rules, within the framework of automatically generating a text analyzer.

Since Liddy fails to disclose or suggest any of the foregoing claim elements, it cannot anticipate nor render obvious claim 12. For at least these reasons, Applicant respectfully requests that the related rejection be withdrawn.

CONCLUSIONS

Applicant's invention is both novel and nonobvious over the prior art for the reasons set forth above. None of the prior art of record, either alone or in combination, teaches each and every element of Applicant's claimed invention.

For all of these reasons, Applicant respectfully asserts that all of claims 1-13 are in condition for allowance. The Examiner's early reconsideration is respectfully requested. If the Examiner has any questions, the Examiner is invited to contact Applicant's attorney at the following address or telephone number:

David Alberti  
c/o Patent Department  
GRAY CARY WARE & FREIDENRICH LLP  
2000 University Avenue  
East Palo Alto, CA 94303-2248  
Telephone: (650) 833-2052

Respectfully submitted,

Gray Cary Ware & Freidenrich LLP



David Alberti  
Reg. No. 43,465

Dated: October 6, 2003

GRAY CARY WARE & FREIDENRICH LLP  
2000 University Avenue  
East Palo Alto, CA 94303-2248  
Telephone: (650) 833-2052  
Facsimile: (650) 833-2001